

OAIster - virtuální katalog OAI zdrojů <http://oaister.umdl.umich.edu/o/oaister>

Charakteristika projektu

Co je OAIster. Je to virtuální katalog vyvinutý uskupením „Digital Library Production“ na Universitě v Michiganu. Tito lidé dostali za úkol vyrobit nástroj, který bude zpřístupňovat vědecky orientované datové zdroje, které byly dříve běžnému uživateli skryté nebo těžce dostupné. Takovéto zdroje nebyly procházeny roboty vyhledávacích společností jako je Google nebo Altavista, které by je následně zobrazily na Internetu. OAIster je dotován grantem nadace Mellone. Stavebním kamenem projektu je využití protokolu OAI, díky němuž jsou zpřístupněny tyto zdroje. Uživatel nezíská pouze jejich metadata, ale zdroje celé. Projekt byl spuštěn v červnu 2002.

Stav

V současné době může návštěvník vyhledávat ve 4,740,751 záznamech jež zpřístupnilo 390 poskytovatelů dat (aktualizováno 3 prosince 2004). Poskytovatel služeb OAIster zpřístupňuje data např. společností NASA (31258 záznamů) nebo rukopisy, fotografie a výtvarné práce z kalifornské oblasti uchovávané v „Archive of California“. V 50.5% zastoupení jsou instituce z U.S.A, 49,5% patří zbývajícím poskytovatelům.

OAI a OAIster

Název OAIster byl odvozen od OAI protokolu. Je to jednoduchý protokol umožňující poskytovatelům služeb automatizované získávání (sklizení) metadat nabízených poskytovateli dat. Díky těmto vlastnostem a díky mnoha volně dostupným softwarovým komponentám je možné protokol snadno implementovat do stávajících systémů digitálních knihoven. Byl navržen v Santa Fe, kam byli svoláni odborníci zabývající se v různých institucích problematikou publikování vědeckých prací v elektronické podobě (pre-prints). Platformu pojmenovali Open Archives Initiative, což je nutno chápat ve smyslu otevřenosti technického řešení skladu elektronicky publikovaných odborných prací. Na to logicky navázala nutnost vzniku služeb, jež zprostředkovaly uživatelům informace o dokumentech uložených v jednotlivých archivech. Identifikuje dvě základní entity: Poskytovatele dat a poskytovatele služeb. Přičemž poskytovatel dat zpřístupňuje prostřednictvím standardizovaného rozhraní metadata, která jsou prostřednictvím tohoto rozhraní sklizena poskytovatelem služeb, který je poskytuje s přidanou hodnotou. Zároveň tato konvence stanovuje požadavek na jednoznačný identifikátor metadatového záznamu na straně poskytovatele dat, definuje základní společné metadatové schéma a rozhraní pro sklizení metadat, založené na podmnožině protokolu Dienst.

Jak už bylo naznačeno, protokol OAI je postaven na čtyřech základních vlastnostech: využití jednoznačných identifikátorů, použití společného metadatového standardu (nekvalifikovaný dublin Core), komunikace prostřednictvím protokolu HTTP a na důsledném využití formátu XML. Společným jmenovatelem těchto vlastností je jednoduchost jak ve smyslu snadnosti implementace, tak i použití protokolu.

Standart nejprve definuje základní pojmy týkající se metadat.

Resource - zdroj, entita která je popisována daným metadatovým objektem.

Repository - server (repositář) provozovaný poskytovatelem dat a poskytující metadatové záznamy prostřednictvím OAI.

Harvester - Klientský program, provozovaný poskytovatelem služeb za účelem získání metadatových záznamů z repositářů.

Item - metadatový objekt uložen v repositáři, z něj se odvozují jednotlivé metadatové záznamy

Identifikátor

Pro umožnění duplikace metadatových záznamů je definován jednoznačný identifikátor metadatového objektu v rámci repositáře. Identifikátor se skládá ze tří částí oddělených dvojtečkami. První část je označení schématu a v případě oai identifikátoru je jím vždy řetězec "oai", druhá část má podobu doménového jména serveru, třetí část pak jednoznačně identifikuje objekt v rámci serveru. Např. oai:HUSeattle.com:3000257

Metadatový záznam

je pro přenos kódovaný v XML. Dělí se na tři části - hlavičku (*header*), vlastní metadatový záznam (*metadata*) a pomocné údaje (*about*)

Skupiny záznamů

Aby nebylo nutné při každém požadavku na sklizení přenášet všechny záznamy, je v OAI protokolu definován mechanismus sdružování metadatových objektů do skupin. Ukazuje se jako výhodný v případě větších repositářů. Každá skupina je definována svým identifikátorem (*setSpec*), krátkým názvem (*setName*) a nepovinným popisem (*setDescription*)

Využití protokolu HTTP

Veškerá komunikace mezi harvesterem a repositářem probíhá prostřednictvím protokolu HTTP. Harvester přitom svoje požadavky pokládá ve formě příkazů GET nebo POST (zakódované jako *application/x-www-form-urlencoded*) a repositář musí podporovat oba tyto způsoby. Celý požadavek pak může vypadat např. takto: `http://arxiv.org/oai2?verb=identify`

Vývoj OAIsteru

Počátkem roku 2001 navrhla knihovna Michiganské Univerzity nadaci Mellone, že za jejich podpory uvede do chodu otevřenou, všeobecně použitelnou službu pro poskytování datových záznamů jako zdroj informací. Byla vytvořena ve spolupráci s Univerzitou v Illinois, která dodala nástroj pro sklizení metadat. Michiganská Univerzita se zaměřila na metadata veřejně dostupných informačních zdrojů. Měly být bez omezeného přístupu s odpovídajícím webovým zastoupením. Služba se snažila obsáhnout co možná největší sbírku zdrojů, která měla být kompletně zpřístupněna internetové obci. Middleware, který Michiganská Univerzita navrhla pro zpřístupnění služby byl volně k dispozici ostatním institucím.

Projekt byl rozdělen do několika fází. První fáze byla čistě výzkumná a uskutečňovala se paralelně s nástrojem na sklizení metadat v Illinois. O projekt se starali dva členové - Kat Hagedorn (knihovnick) a Mike Burek (programátor).

Hageborn začal kontaktovat potenciální zdroje a Burek se seznamoval s charakterem OAI.

V druhé fázi se uskutečňovaly následující kroky

1) Shromáždění jednotlivých záznamů poskytnutých harvestrem. Metadata byla kódována v DC a byla sklížena pomocí XML. Tyto záznamy byly slučovány do obsáhlých souborů s adresářovou strukturou.

2) Vyjmutí nedigitálních objektů.

3) Znormalizování hlaviček.

4) Přidání informace o instituci do každé položky.

5) Upravení UTF-8 na ISO8859-1 kvůli indexování a vyhledávání.

6) Převedení DC záznamu do místního formátu BibClass přes XSLT.

Převodní nástroj byl vynalezen na Michiganské Univerzitě. Je napsán v Javě a opírá se o technologii XSLT.

7) Indexování a předkládání těchto metadat skrz rozhraní používající XPAT Search engine, který byl získán přes Digital Library eXtension Service (DLXS).

8) Sečtení záznamů.

V polovině roku 2002 Michigan Univerzity uvolnila svoji službu zvanou OAIster a začala s distribucí middleware pro OAIster přes DLXS. Následovaly fáze hodnocení, experimentování, revize a rozšíření.

Vyhledávání

Máme dvě možnosti vyhledávání. Buď prohledáváním všech polí popisujících digitální zdroj nebo zadáním řetězce do specifických polí -titul, autor, předmět a typ zdroje. Každý nalezený záznam obsahuje pouze ty pole, které jsou v něm přítomny. Po nalezení záznamu se přes kliknutí na link přeneseme na požadovaný zdroj.

Závěr

Význam takovýchto projektů spočívá v systematickém zpřístupňování velkého množství informací na Internetu. Uživatel má velmi jednoduchý přístup k informacím z mnoha různých tematických oblastí. Poskytovatele dat v relativně snadném zapsání repositáře k poskytovateli služby. Uživatelské prostředí OAIstru je graficky příjemné, jednoduché. Ve vyhledávání se mi líbí rozdělení výsledků hledání dle institucí.

Jedním z důležitých předpokladů pro rozšíření podobných projektů je dostupnost a kvalita obsahu, který by měl být prostřednictvím těchto projektů nabízen. Věřím, že jakmile se objeví dostatečné množství repositářů s dostatečně zajímavým obsahem, začnou se více objevovat kvalitní služby, které tento obsah zpřístupní ať už přímo, nebo prostřednictvím informačních bran.

Použitá literatura a WWW odkazy

<http://oaister.umdl.umich.edu/o/oaister/>

http://www.openarchives.org/meetings/SantaFe1999/sfc_entry.htm

<http://www.openarchives.org>

http://knihovny.cvut.cz/akp2003/sbornik/05_zabicka.pdf

<http://dlxs.org/about/prospectus.html>

<http://www.cvt.dk/oaisem/docs/Kat/dtu.pdf>